
自动化学报
Acta Automatica Sinica
ISSN 0254-4156,CN 11-2109/TP

《自动化学报》网络首发论文

题目：　性能函数引导的无人机集群深度强化学习控制方法
作者：　王耀南，华和安，张辉，钟杭，樊叶心，梁鸿涛，常浩，方勇纯
DOI：　10.16383/j.aas.c240519
收稿日期：　2024-07-23
网络首发日期：　2025-01-03
引用格式：　王耀南，华和安，张辉，钟杭，樊叶心，梁鸿涛，常浩，方勇纯．性能函数引导的无人机集群深度强化学习控制方法[J/OL]．自动化学报．https://doi.org/10.16383/j.aas.c240519



第 xx 卷　第 x 期　自 动 化 学 报　Vol. xx, No. x
xxxx 年 xx 月　ACTA AUTOMATICA SINICA　xxxx, xxxx

性能函数引导的无人机集群深度强化学习控制方法

王耀南[1,2]　华和安[1,2]　张 辉[1,3]　钟 杭[1,3]　樊叶心[1,3]　梁鸿涛[1,2]　常 浩[1,2]　方勇纯[4,5]

摘　要　针对无人机集群系统, 提出了一种性能函数引导的深度强化学习控制方法, 同时评估性能函数的示范经验与学习策略的探索动作, 保证了高效可靠的策略更新, 实现了无人机集群系统的高性能控制. 首先, 利用领航-跟随集群框架, 将无人机集群的控制问题转化为领航-跟随框架下的跟踪问题, 进而提出了基于模型的跟踪控制方法, 利用性能函数将集群编队误差约束在给定范围内, 实现了无人机集群的模型驱动控制. 接下来, 为了解决复杂工况下性能函数极易失效难题, 将深度强化学习方法和性能函数驱动方法结合, 提出了性能函数引导的深度强化学习控制方法, 利用性能函数的示范经验辅助训练强化学习网络, 通过同时评估探索与示范动作, 保证学习策略显著优于性能函数驱动控制方法, 有效提高了无人机编队控制精度与鲁棒性. 实验结果表明, 该方法能够显著提升无人机集群的控制精度, 实现了兼顾鲁棒性与飞行精度的高性能集群控制.

关键词　无人机集群, 深度强化学习, 引导式学习, 智能编队控制

引用格式　王耀南, 华和安, 张辉, 钟杭, 樊叶心, 梁鸿涛, 常浩, 方勇纯. 性能函数引导的无人机集群深度强化学习控制方法. 自动化学报, xxxx, **xx**(x): x−xx

DOI　10.16383/j.aas.c240519　**CSTR**　32138.14.j.aas.c240519

Performance Function-Guided Deep Reinforcement Learning Control for UAV Swarm

WANG Yao-Nan[1,2]　HUA He-An[1,2]　ZHANG Hui[1,3]　ZHONG Hang[1,3]　FAN Ye-Xin[1,3]
LIANG Hong-Tao[1,2]　CHANG Hao[1,2]　FANG Yong-Chun[4,5]

Abstract　In this paper, a novel performance function-guided deep reinforcement learning control method is proposed for the UAV swarm system, where the demonstration of performance function and deep reinforcement learning exploration are both utilized to guarantee efficient and reliable policy updating, achieving high-performance control of the UAV swarm system. Firstly, based on the leader-follower framework, the UAV swarm control problme is formulated as the leader follower tracking problem, and then, the model-based tracking control is proposed, where the performance function is deigsned to constrain the tracking error within a given range, thereby achieving UAV formation. Then, to address the invalid problem of performance function under complex working conditions, the deep reinforcement learning and the performance function-driven methods are combined to propose the performance-function-guided deep reinforcement learning control method, where the demonstration of performance function is used to train the reinforcement learning network, and the learning strategy is significantly better than the performance function-driven control method by simultaneously evaluating exploration and demonstration actions, thereby improving the accuracy and robustness of UAV formation control effectively. Comparative experimental results show that the proposed method significantly improves the control performace of UAV swarms, realizing high-performance swarm control with both robustness and flight accuracy.

Key words　UAV swarm, deep reinforcement learning, guided learning, intelligent formation control

Citation　Wang Yao-Nan, Hua He-An, Zhang Hui, Zhong Hang, Fan Ye-Xin, Liang Hong-Tao, Chang Hao, Fang Yong-Chun. Performance function-guided deep reinforcement learning control for UAV swarm. *Acta Automatica Sinica*, xxxx, **xx**(x): x−xx

随着机器人与人工智能技术的发展, 无人机以其低速飞行、机动性强、灵活性高等优势得到广泛应用[1−3]. 无人机集群的可扩展性高, 可执行大规模

收稿日期 2024-07-23　录用日期 2024-11-11
Manuscript received July 23, 2024; accepted November 11, 2024
科技创新 2030-“新一代人工智能”重大项目 (2021ZD0114503), 国家自然科学基金 (62403190, 62427813, 62433010) 资助
Supported by the National Key Research and Development Program of China (2021ZD0114503), National Natural Science Foundation of China (62403190, 62427813, 62433010)
本文责任编委
Recommended by Associate Editor
1. 机器人视觉感知与控制技术国家工程研究中心 长沙 410082　2. 湖南大学电气与信息工程学院 长沙 410082　3. 湖南大学机器人学院 长沙 410082　4. 南开大学机器人与信息自动化研究所 天津 300350　5. 南开大学智能技术与机器人系统研究院 深圳 518083
1. National Engineering Research Center of Robot Visual Perception and Control Technology, Changsha 410082　2. College of Electrical and Information Engineering, Hunan University, Changsha 410082　3. School of Robotics, Hunan University, Changsha 410082　4. Institute of Robotics and Automatic Information Systems, Nankai University, Tianjin 300350　5. Institute of Intelligence Technology and Robotic Systems, Nankai University, Shenzhen 518083

复杂任务, 相关研究受到了学术界和产业界的广泛关注[4−6]. 为了安全高效精准地完成既定任务, 集群控制算法至关重要. 然而, 考虑到无人机的非线性特性、开环不稳定和欠驱动特性, 设计有效的集群控制算法极具挑战. 尤其是, 无人机集群的任务复杂多变, 面临动态场景、复杂干扰等极端情况[7]. 因此, 设计智能高效的集群控制策略, 保证无人机间的高效协作, 具有重大的理论和实践意义. 为此, 国内外研究人员已经提出了许多标志性的控制方法[8−11], 包括基于行为的集群控制、虚拟结构法、领航-跟随法、人工势场法、一致性方法等, 实现了有效的无人机集群控制, 其中的领航-跟随法以其出色的集群控制精度、策略扩展性等优势得到了广泛研究.

为了实现精准的领航-跟随集群控制, 研究人员提出了许多可行的模型驱动控制策略[12−14], 其核心思想是充分利用无人机集群系统的动力学模型, 并在此基础上设计领航-跟随控制方法, 并对设计的闭环系统进行稳定性分析. 例如, 文献 [14] 提出了一种分布式领航-跟随控制算法, 并基于 Lyapunov 理论证明了闭环系统的渐近稳定性. 文献 [15] 考虑了无人机集群模型中的非线性特性、参数扰动、通信延迟、外部干扰等因素, 设计位置与姿态控制器, 实现了无人机集群的鲁棒编队. 进而, 文献 [16] 针对无人机集群中的时变通讯延迟, 设计了分布式编队控制协议, 证明了编队误差将在有限时间内收敛. 此外, 为了进一步约束集群控制误差, 文献 [17] 通过构造编队误差边界, 提出了一种预设性能的自适应编队控制方法, 通过设计障碍 Lyapunov 函数, 将集群系统状态始终约束在预定范围内. 尽管上述模型驱动的集群控制方法实现了无人机的编队控制, 但是面向实际应用, 仍有许多关键问题亟待解决. 首先, 尽管利用无人机集群系统的动力学模型, 可以设计闭环系统稳定的集群编队控制方法, 但是集群控制中的一些关键指标仍没有保证, 特别是集群的控制误差可能超出安全范围, 引发事故. 其次, 基于障碍函数的控制方法能够将集群误差抑制在给定范围内, 但是在逼近约束边界时其控制输出面临饱和风险, 存在控制失效、闭环系统崩溃等安全隐患.

随着人工智能技术的发展, 学习驱动的机器人技术研究受到越来越多研究人员的关注[18−21]. 不同于模型驱动的集群控制方法, 深度学习驱动的控制设计利用海量数据训练得到的控制策略, 具有很强的适应性和灵活性[22−24]. 其中, 深度强化学习通过与环境的试错交互, 利用最大化奖励函数寻找最优控制策略, 已经成功应用于多种机器人控制. 例如, 文献 [25] 提出了一种旋翼无人机深度强化学习控制方法, 通过设计学习网络, 以无人机状态作为输入, 直接控制无人机驱动器, 实现高效控制. 文献 [26] 将注意力机制引入深度强化学习, 解决了多智能体高效合作问题. 利用好奇心机制, 文献 [27] 提出了一种好奇心驱动的深度强化学习控制方法, 实现了无人机机动飞行. 此外, 通过引入积分补偿输入, 文献 [28] 设计了深度强化学习无人机控制方法, 提高了稳态控制精度. 文献 [29] 采用深度强化学习与领航-跟随框架相结合的设计方法, 提出了一种编队控制方法, 实现无人机集群高效编队. 为了进一步提高无人机学习控制的可靠性, 通过引入模型的先验信息设计的深度强化学习控制策略能兼顾学习效率与控制精度. 例如, 文献 [30] 设计的混合深度强化学习控制方法, 将基于模型的设计与学习策略线性组合, 以减少复杂的不确定因素对无人机的影响. 此外, 文献 [31] 提出了一种模型预测控制与深度强化学习相结合的控制方法, 利用学习策略补偿未建模动态. 尽管上述学习驱动的无人机控制方法实现了无人机的飞行控制, 但是仍有许多关键问题亟待解决. 一方面, 深度强化学习通过与环境的试错交互, 学习到收敛的控制策略. 但是, 无人机集群的复杂非线性动力学, 加之执行任务时面临的动态场景、复杂干扰等极端情况, 深度强化学习策略难以同时应对上述挑战, 往往无法探索到安全且可靠的集群控制策略. 另一方面, 几乎所有的深度强化学习算法都没有考虑到集群控制中的误差约束等关键指标, 在训练过程中没有充分利用无人机集群系统的动力学模型先验信息, 导致得到的学习控制策略没有安全保障, 极端环境中极易失效.

针对上述问题, 本文提出了性能函数引导的深度强化学习无人机集群控制方法. 具体而言, 通过构造无人机集群编队误差边界, 设计性能函数将系统状态约束在预定边界内. 然后, 设计了双 critic 架构的深度强化学习网络架构, 并引入性能函数的示范经验, 通过同时评价学习策略的随机探索动作与性能函数的示范输出, 实现对探索动作的精准判断. 在此基础上, 使用显著优于示范经验的探索动作更新策略, 有效提高了无人机编队控制精度与鲁棒性. 最后, 通过集群飞行实验验证了所提方法的有效性. 本文的主要贡献总结如下:

1) 针对无人机集群系统, 提出了一种性能函数引导的深度强化学习控制方法, 提高了集群系统的飞行控制精度与鲁棒性.

2) 使用性能函数的示范经验, 引导深度强化学习策略探索更好的控制策略, 对训练过程中的探索动作实现准确评价.

3) 实验结果表明所提的性能函数引导的深度强化学习控制方法能够实现准确的动作评价、高效的策略更新和精准的集群控制.

本文接下来的内容如下:在第 1 节中介绍无人机动力学模型, 并且介绍领航-跟随编队控制框架; 在第 2 节中设计性能函数驱动的编队控制引导策略, 提出性能函数引导的控制算法, 并在此基础上针对提出的引导深度强化策略, 设计训练-评估算法; 在第 3 节中设计集群飞行的仿真实验, 验证方法的可行性和有效性; 最后, 在第 4 节中总结本文内容.

1　问题提出

1.1　无人机动力学模型

本文考虑由 n 个无人机组成的集群系统, 其中第 i 个无人机的动力学模型表示如下 [22]:

$$\begin{aligned}&\dot{\boldsymbol{p}}_i=\boldsymbol{v}_i\\&m_i\dot{\boldsymbol{v}}_i+m_ig\boldsymbol{e}_3=f_iR_i\boldsymbol{e}_3\\&\dot{R}_i=R_i\hat{\boldsymbol{\Omega}}_i\\&\mathbb{I}_i\dot{\boldsymbol{\Omega}}_i+\boldsymbol{\Omega}_i\times\mathbb{I}_i\boldsymbol{\Omega}_i=\boldsymbol{\tau}_i\end{aligned}\tag{1}$$

其中 $\boldsymbol{p}_i=[x_i\ y_i\ z_i]^\top\in\mathbb{R}^3$, $\boldsymbol{v}_i=[v_{ix}\ v_{iy}\ v_{iz}]^\top\in\mathbb{R}^3$ 分别表示无人机 i 在惯性系中的位置与速度; m 表示无人机 i 的质量; $g\in\mathbb{R}$, $\boldsymbol{e}_3=[0\ 0\ 1]^\top\in\mathbb{R}^3$ 分别为重力加速度常数和竖直向上方向单位向量; $\mathbb{I}_i\in\mathbb{R}^{3\times3}$ 表示无人机 i 的转动惯量; $R_i\in\mathbb{R}^{3\times3}$, $\boldsymbol{\Omega}_i\in\mathbb{R}^3$ 表示无人机 i 从连体系到惯性系的旋转矩阵以及其在连体系中的旋转角速度; $f_i\in\mathbb{R}$ 表示无人机 i 的升力; $\boldsymbol{\tau}_i\in\mathbb{R}^3$ 表示无人机 i 的转动力矩; $\hat{}$ 表示 $\mathbb{R}^3\to\mathrm{SO}(3)$ 映射, 将向量变换为反对称矩阵, $^\vee$ 表示 $\mathrm{SO}(3)\to\mathbb{R}^3$ 其反变换, 将反对称矩阵转化为相应的向量.

1.2　无人机领航-跟随编队控制

无人机集群中的领航-跟随编队控制问题是控制 n 个无人机系统保持一定的队形并同时跟踪给定飞行轨迹. 领航-跟随编队中只有领航无人机能够获得飞行轨迹信息, 并且每个无人机只从其直接领航无人机节点获得信息. 换言之, 领航-跟随编队的通信拓扑结构为有向生成树, 如图 1 所示. 具体而言, 首先, 引入虚拟领航无人机 U_0, 并根据任务设计其参考轨迹 $\boldsymbol{p}_0=[x_0,\ y_0,\ z_0]^\top\in\mathbb{R}^3$. 虚拟领航无人机 U_0 与无人机 U_1 组成一对领航-跟随编队, U_1 从 U_0 处接收参考轨迹. 接下来, 任意无人机节点 U_i 将仅从其直接领航无人机 U_{i-1} 处获得状态向量信息. 此时, 虚拟领航无人机确定整个集群系统的运动轨迹,

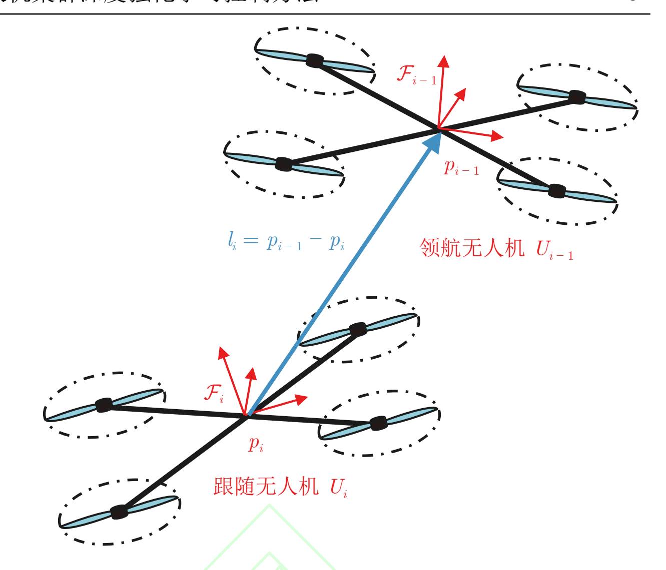


图 1　无人机领航-跟随编队模型示意图

Fig.1　Schematic diagram of drone pilot-following formation model

其余 n 个无人机跟踪其给定的领航者, 形成 n 对领航-跟随结构.

在此基础上, 将编队问题转换成局部轨迹跟踪问题, 如图 1 所示. 无人机 U_i 从领航无人机 U_{i-1} 处获得其状态向量信息, 计算当前编队方位信息 $\boldsymbol{l}_i=\boldsymbol{p}_{i-1}-\boldsymbol{p}_i$, 考虑期望编队方位为 $\boldsymbol{l}_{di}=[\Delta x_{di}\ \Delta y_{di}\ \Delta z_{di}]^\top\in\mathbb{R}^3$,相对误差 $\boldsymbol{e}_{pi}=\boldsymbol{l}_{di}-\boldsymbol{l}_i=[e_{pix}\ e_{piy}\ e_{piz}]^\top\in\mathbb{R}^3$ 定义如下

$$\begin{aligned}&e_{pix}=x_i-x_{i-1}+\Delta x_{di},\\&e_{piy}=y_i-y_{i-1}+\Delta y_{di},\\&e_{piz}=z_i-z_{i-1}+\Delta z_{di}\end{aligned}\tag{2}$$

接下来, 定义 $\boldsymbol{p}_{di}=[x_{i-1}-\Delta x_{di}\ \ y_{i-1}-\Delta y_{di}\ \ z_{i-1}-\Delta z_{di}]^\top\in\mathbb{R}^3$, 则 $\boldsymbol{e}_{pi}=\boldsymbol{p}_i-\boldsymbol{p}_{di}$. 针对第 i 个无人机, 构造 x, y, z 方向上的飞行误差约束如下:

$$-\beta_{ij}<e_{pij}<\beta_{ij},\quad i=0,\ \cdots,\ n,\quad j=x,\ y,\ z\tag{3}$$

其中 $\beta_{ij}>0$ 表示误差约束, 用以限制无人机 i 在 x, y, z 方向上的编队飞行误差. 进而, 提出本文的集群控制目标如下, 设计集群中每个无人机的反馈控制律 f_i 和 $\boldsymbol{\tau}_i$, 使得:

$$\lim_{t\to\infty}\boldsymbol{e}_{pi}=0\tag{4}$$

且无人机集群始终满足约束式 (3).

2　主要内容

本节提出了一种性能函数引导的深度强化学习集群控制方法, 如图 2 所示. 首先, 基于无人机系统 (1) 的分层特性, 分别构造内环与外环子系统. 进

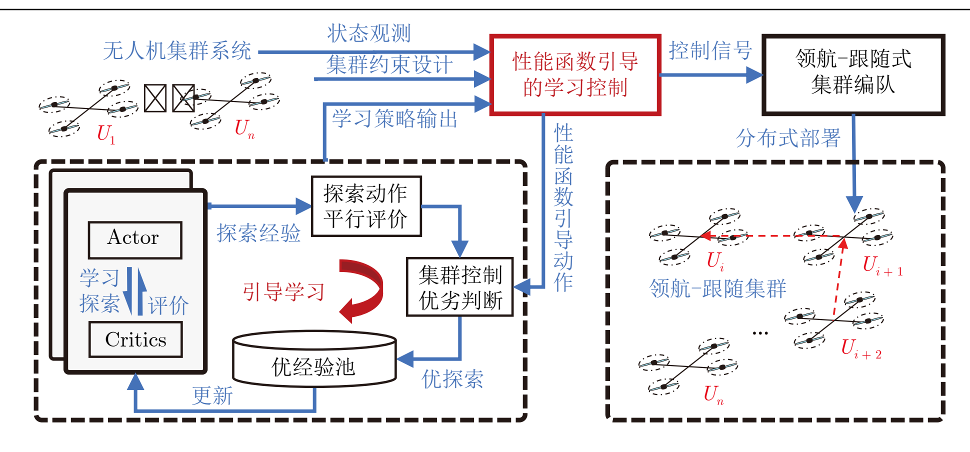


图 2 性能函数引导的深度强化学习集群控制框架

Fig. 2 Performance-driven cluster control bootstrap policy framework

而, 针对集群控制目标 (3)、(4), 设计性能函数驱动的集群控制引导策略. 其中, 通过引入辅助变量与虚拟控制输入, 利用性能函数将集群控制误差约束在给定范围内 (3). 在此基础上, 提出性能函数引导的深度强化学习控制方法, 解决复杂工况下性能函数极易失效难题.

2.1 性能函数驱动的集群控制引导策略设计

本小节将无人机集群的控制问题转化为领航-跟随框架下的跟踪问题, 并提出基于模型的控制方法. 首先, 针对无人机 i 外环位置与内环姿态子系统, 引入辅助变量描述飞行的位姿与姿态误差; 其次, 考虑到无人机的欠驱动特性与变量间耦合关系, 设计虚拟控制变量 $\boldsymbol{F}_i \in \mathbb{R}^3$ 与耦合变量 $\boldsymbol{d}_i \in \mathbb{R}^3$ 将系统转化为级联结构; 在上述基础上, 计算得到了无人机 i 的开环动力学方程; 最后, 基于飞行误差约束与系统开环动力学方程, 设计性能函数驱动的领航-跟随控制方法, 利用性能函数进行反馈调节, 实现集群控制.

首先, 针对无人机 i, 构造外环位置子系统中的辅助变量 $\boldsymbol{\zeta}_i = [\zeta_{ix} \quad \zeta_{iy} \quad \zeta_{iz}]^\top \in \mathbb{R}^3$ 与内环姿态子系统中的辅助变量 $\boldsymbol{\eta}_i = [\eta_{ix} \quad \eta_{iy} \quad \eta_{iz}]^\top \in \mathbb{R}^3$ 如下:

$$\boldsymbol{\zeta}_i = m_i \boldsymbol{e}_{vi} + K_{pi} \boldsymbol{e}_{pi} \tag{5}$$

$$\boldsymbol{\eta}_i = \boldsymbol{e}_{\Omega i} + k_{Ri} \boldsymbol{e}_{Ri} \tag{6}$$

其中 $K_{pi} = \mathrm{diag}[K_{pix},\ K_{piy},\ K_{piz}] \in \mathbb{R}^{3\times 3}$ 为正定的增益矩阵, $k_{Ri} \in \mathbb{R}$ 表示正的增益参数. $\boldsymbol{e}_{vi} = \dot{\boldsymbol{e}}_{pi} \in \mathbb{R}^3$ $\boldsymbol{e}_{\Omega i} = \boldsymbol{\Omega}_i - R_i^\top R_{di} \boldsymbol{\Omega}_{di}$ $\boldsymbol{e}_{Ri} = \frac{1}{2}(R_{di}^\top R_i - R_i^\top R_{di})^\vee$ 分别为第 i 个无人机的速度误差、角速度误差与姿态误差.

其次, 由于旋翼无人机的欠驱动特性, 外环位置子系统输入 f_i 和内环姿态子系统状态 R_i 相互耦合, 难以直接设计飞行控制算法. 为此, 引入虚拟控制输入变量 $\boldsymbol{F}_i \in \mathbb{R}^3$ 与耦合变量 $\boldsymbol{d}_i \in \mathbb{R}^3$, 将系统转换为级联结构. 具体而言, 第 i 个旋翼无人机系统的输入为 $f_i R_i \boldsymbol{e}_3$ 等效表示为:

$$f_i R_i \boldsymbol{e}_3 = \boldsymbol{F}_i + \boldsymbol{d}_i \tag{7}$$

其中虚拟输入 $\boldsymbol{F}_i$ 的表达式如下

$$\boldsymbol{F}_i = \frac{f_i R_{di} \boldsymbol{e}_3}{\boldsymbol{e}_3^\top R_{di}^\top R_i \boldsymbol{e}_3} \tag{8}$$

值得一提的是, 式 (8) 揭示了虚拟控制输入 $\boldsymbol{F}_i$ 和输入推力 f_i 之间的关系, 其值可以任意设计. $\boldsymbol{d}_i$ 表示无人机外环位置子系统和内环姿态子系统之间的耦合变量[22], 可表示为 $\boldsymbol{d}_i = \frac{1}{\boldsymbol{e}_3^\top R_{di}^\top R_i \boldsymbol{e}_3}[f_i(\boldsymbol{e}_3^\top R_{di}^\top \times R_i \boldsymbol{e}_3) R_i \boldsymbol{e}_3 - f_i R_{di} \boldsymbol{e}_3]$. 在此基础上, 内环姿态子系统中的期望姿态表示为 $R_{di} = [\boldsymbol{b}_{2di} \times \boldsymbol{b}_{3di} \quad \boldsymbol{b}_{2di} \quad \boldsymbol{b}_{3di}]$, 其中 $\boldsymbol{b}_{2di}$ 的值由式 $\boldsymbol{b}_{2di} = \boldsymbol{b}_{3di} \times \boldsymbol{b}_{1di} / \|\boldsymbol{b}_{3di} \times \boldsymbol{b}_{1di}\|$ 计算, 且 $\boldsymbol{b}_{1di}$ 不与 $\boldsymbol{b}_{3di}$ 平行. 然后, 根据虚拟控制输入的值, 计算得到 $\boldsymbol{b}_{3di} = \boldsymbol{F}_i / \|\boldsymbol{F}_i\|$, 其中 $\|\boldsymbol{F}_i\| \neq 0$.

在此基础上, 根据系统的动力学模型 (1), 无人机 i 期望的角速度可表示为 $\boldsymbol{\Omega}_{di} = (R_{di}^{-1} \dot{R}_{di})^\vee$. 进而, 基于无人机 i 动力学模型 (1)、虚拟控制输入 (7), 得到如下开环动力学方程:

$$\dot{\boldsymbol{\zeta}}_i = -m_i g \boldsymbol{e}_3 - m_i \ddot{\boldsymbol{p}}_{di} + K_{pi} \boldsymbol{e}_{vi} + \boldsymbol{F}_i + \boldsymbol{d}_i \tag{9}$$

$$\mathbb{I}_i \dot{\boldsymbol{\eta}}_i = \boldsymbol{\tau}_i + \boldsymbol{\tau}_{fi} \tag{10}$$

其中 $\boldsymbol{\tau}_{fi}$ 表示内环姿态子系统的前馈项, $\boldsymbol{\tau}_{fi} = -\boldsymbol{\Omega}_i \times \mathbb{I}_i \boldsymbol{\Omega}_i + \mathbb{I}_i(\hat{\boldsymbol{\Omega}}_i R_i^\top R_{di} \boldsymbol{\Omega}_{di} - R_i^\top R_{di} \dot{\boldsymbol{\Omega}}_{di}) + k_{\eta i} \mathbb{I}_i(\mathrm{tr}[R_i^\top \times R_{di}] I - R_i^\top R_{di}) \boldsymbol{e}_{\Omega i}/2$.

最后, 根据飞行误差约束 (3) 以及系统的开环动力学方程 (9)、(10), 设计性能函数驱动的领航-跟随控制方法如下:

$$\boldsymbol{F}_i = m_i g \boldsymbol{e}_3 + m_i \ddot{\boldsymbol{p}}_{di} - K_{pi} \boldsymbol{e}_{vi} - K_{\zeta i} \boldsymbol{\zeta}_i - \Lambda_i \boldsymbol{e}_{pi} \tag{11}$$

$$\boldsymbol{\tau}_i = -k_{\eta i}\boldsymbol{\eta}_i - k_i\boldsymbol{e}_{Ri} - \boldsymbol{\tau}_{fi} \tag{12}$$

其中 $K_{\zeta i} \in \mathbb{R}^{3\times3}$ 表示正定增益矩阵, $k_{\eta i}$, $k_i \in \mathbb{R}^+$ 为正的增益参数. 定义 $\Lambda_i = \mathrm{diag}[\varphi_{ix}, \varphi_{iy}, \varphi_{iz}] \in \mathbb{R}^{3\times3}$ 为性能函数, 其值如下:

$$\varphi_{ij} = \frac{k_{\varphi ij}\beta_{ij}^2}{(\beta_{ij}^2 - e_{pij}^2)^2}, \quad i = 0, \cdots, n, \quad j = x, y, z \tag{13}$$

其中, φ_{ij} 第 i 个无人机在 j 轴上的性能函数, $k_{\varphi ij}$ 表示正的控制参数. 定义 $\boldsymbol{\varphi}_i = [\varphi_{ix}\ \ \varphi_{iy}\ \ \varphi_{iz}]^\top \in \mathbb{R}^3$ 为性能函数的向量形式. 接下来, 将性能函数驱动的控制 (11), (12) 代入系统的开环动力学方程, 得到闭环系统如下:

$$\dot{\boldsymbol{\zeta}}_i = -K_{\zeta i}\boldsymbol{\zeta}_i - \Lambda_i\boldsymbol{e}_{pi} + \boldsymbol{d}_i \tag{14}$$

$$\mathbb{I}_i\dot{\boldsymbol{\eta}}_i = -k_{\eta i}\boldsymbol{\eta}_i - k_i\boldsymbol{e}_{Ri} \tag{15}$$

本文通过引入入辅助函数和虚拟控制输入, 将无人机 i 的动力学模型转换为级联架构. 进而, 设计性能函数并提出模型驱动的集群控制方法 (11)、(12). 通过对集群误差的实时观测, 利用性能函数动态地调整控制作用, 成功将误差约束在预设范围内, 如图 2 所示. 具体而言, 每当集群飞行误差趋近于预设边界时, 性能函数将随之变化, 通过主动的反馈调节, 将系统状态拉回预设范围内.

2.2　引导深度强化学习策略设计

尽管所提算法 (11)、(12) 能够驱动无人机集群中每个领航-跟随编队完成精准飞行任务, 并且能将其编队误差约束在给定区间 (3). 然而, 性能函数的约束项可能引起系统输入饱和.特别是, 在无人机执行任务过程中, 将不可避免的出现外界干扰、自身模型漂变等意外情况, 此时所提控制算法 (11)、(12) 可能失效. 一方面, 所提性能函数能有效约束飞行误差的必要条件是系统输入可任意设计. 实际上, 由于无人机的物理约束, 系统的实际控制输入存在上界 $F_{i,max}$, 在不确定干扰因素的作用下, 误差 $\boldsymbol{e}_{pij}$ 可能接近性能函数边界 β_{ij}, 性能函数 Λ_i 将如图 2 所示快速增大, 由公式 (11) 可知控制输入 F_i 也将快速增大, 极易造成驱动器饱和, 进而导致系统误差越过性能函数边界、控制失效等严重后果. 另一方面, 所提性能函数仅能约束误差边界 β_{ij} 内的状态. 在不确定干扰因素作用下, 系统误差一旦越过边界, 性能函数 Λ_i 将立即失效, 闭环系统发散, 将不可避免地导致安全事故. 为此, 本文提出性能函数引导的深度强化学习控制方法, 充分利用性能函数的优点, 同时解决系统存在的控制失效问题.

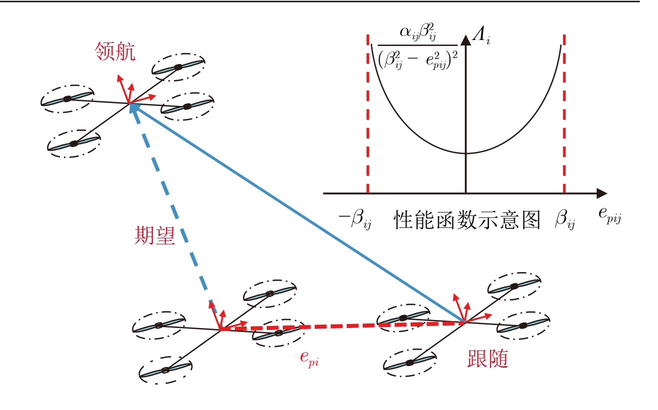


图 3　性能函数驱动的集群控制引导策略框架

Fig.3　Performance function driven cluster control guidance policy framework

本文通过充分利用性能函数的示范经验与强化学习的随机探索, 提出性能函数引导的深度强化学习控制框架. 具体而言, 采用演员-评论家 (actor-critic) 框架构建深度强化学习控制策略, 其中, 系统的状态变量 $\boldsymbol{s}_i = [\boldsymbol{e}_{pi}^\top, \boldsymbol{e}_{vi}^\top]^\top \in \mathbb{R}^6$ 为强化学习策略输入, $\boldsymbol{a}_i = \Lambda_i$ 为策略输出, 其中使用随机探索策略实现迭代寻优, 探索策略满足 $\mathcal{N}(0, \sigma_1)$ 分布. 接下来, 针对高精度编队控制任务目标, 设计策略的奖励函数为 $r_i = -\|\boldsymbol{e}_{pi}\|^2$, 该奖励函数能够寻找到最小化编队误差的控制策略, 实现高精度集群飞行的控制目标. 值得一提的是, 大多数发表的深度强化学习方法更新都采用随机探索策略, 根据 critic 的在线评估对随机探索到的状态动作映射, 寻找高回报策略. 但是, 这些方法难以直接应用于无人机集群控制. 首先, 大多数已发表深度强化学习方法仅适用于低维空间系统, 如倒立摆系统多数为单自由度或双自由度.相比之下, 无人机有 6 个自由度, 其控制难度呈指数上升. 其次, 现有深度强化学习控制方法没有充分利用系统的先验知识, 特别是无人机系统中复杂的状态耦合使得传统的随机探索方法难以获得可靠结果. 为此, 本文提出的性能函数引导的深度强化学习旨在保证高效的策略更新. 具体而言, 采用性能函数驱动的编队控制 (11) 辅助深度强化学习进行高效的探索, 保存优秀的探索经验用于策略更新. 这样一来, 性能函数将驱动深度强化学习算法不断探索更好的策略, 最终得到优于基准性能函数的控制策略, 如 (16) 所示.

$$Q_k = V_k(\boldsymbol{s}_i, \boldsymbol{a}_i) - V_k(\boldsymbol{s}_i, \boldsymbol{\varphi}_i), \ k = 1, 2 \tag{16}$$

在此基础上, 计算双 critic 中的保守估计: $Q = \min\{Q_1, Q_2\}$, 保存 $Q > 0$ 的探索经验用于策略更新. 首先, 引入性能函数基准示范经验 $\boldsymbol{\varphi}_i$, 其中 $\varphi_{ij} = k_{\varphi ij}\beta_{ij}^2/(\beta_{ij}^2 - e_{pij}^2)^2$, $j = x, y, z$. 其次, 设计双

critic 评价策略, 即通过两个独立的 critic 评价, 使用其中的保守估计评价 actor 策略. 具体而言, 首先计算第 k 个评论家的评价值 $V_k(\boldsymbol{s}_i,\ \boldsymbol{a}_i)$ 和 $V_k(\boldsymbol{s}_i,\ \boldsymbol{\varphi}_i)$, 其中 $V_k(\boldsymbol{s}_i,\ \boldsymbol{a}_i)$ 表示对探索动作 $\boldsymbol{a}_i$ 的评价值, $V_k(\boldsymbol{s}_i,\ \boldsymbol{\varphi}_i)$ 是 $\boldsymbol{\varphi}$ 性能函数输出的评价. 通过同时评价探索动作与性能函数策略, 可以有效分辨当前探索到策略的优劣.

所提算法的训练如算法 1 所示. 首先, 由 α_1, α_2 和 α_3 分别参数化 actor 网络 π, critic 网络 V_1 和 V_2. 在每个回合的训练开始前将初始化系统状态. 在训练中, 智能体的状态为 $\boldsymbol{s}_i$, 智能体根据策略 π 选择执行动作 $\boldsymbol{a}_i$, 之后从环境获得奖励 r_i, 并进入下一个状态 $\boldsymbol{s}_{\text{next}}$. 具体而言, 在每一步的迭代中, 智能体将同时评价探索动作与性能函数策略, 利用 (16) 评估每次探索的质量并给探索经验分级. 若是 $Q > 0$, 则认为当前探索策略优于基准策略; 若是 $Q \le 0$, 则认为当前探索策略较差. 在评价中使用两个独立的 critic 决策中的保守值, 进一步保证了对优秀探索动作的准确判断. 通过这种方式, 可以筛选出显著优的学习经验, 用以更新策略, 进而保证高效可靠的策略迭代. 此外, 在策略部署阶段, 采用性能函数的基线动作评价策略动作的优劣, 通过 Q 值的分布来评估所提方法的泛化能力.

算法 1. 性能函数引导的集群控制训练算法

随机初始化网络 π 、V_1 、V_2 的参数 α_k, 并以相同参数初始化目标网络 $\alpha'_k \leftarrow \alpha_k$, $k = 1,\ 2,\ 3$
初始化经验池 $\mathcal{B}_1$ 和 $\mathcal{B}_2$
for $m \le M_{\max}$ do
  初始化无人机 i 的状态
  观测其初始状态 $\boldsymbol{s}_i$
  for $n \le N_{\max}$ do
    根据当前状态计算性能函数输出 $\boldsymbol{\varphi}_i$
    随机探索当前学习策略输出 $\boldsymbol{a}_i$
    输出学习控制信号 $\boldsymbol{a}_i$, 观测状态 $\boldsymbol{s}_{\text{next}}$
    计算当前动作奖励 r_i
    保存当前的交互数据到 $\mathcal{B}_1$
    同时评价当前动作与引导动作的 Q 值
    if $Q > 0$ then
      保存交互数据到 $\mathcal{B}_2$
    end
    if 经验池 $\mathcal{B}_2$ 已满 then
      采样 $\mathcal{B}_2$ 中的数据更新
    else
      采样 $\mathcal{B}_1$ 中的数据更新
    end
  end
end

为了高效地利用交互经验, 本文采用经验回放技术减少训练数据之间的相关性, 同时构造双经验池 $\mathcal{B}_1$ 和 $\mathcal{B}_2$, 分别存储所有的交互经验和显著优于基准策略的高价值的经验. 具体而言, 经验池大小固定, 数据遵从先入先出原则, 当有新的数据填充后, 末尾的数据将被淘汰. 由于策略采用随机初始化, 在训练开始时只有少量的高价值的经验, 策略将从 $\mathcal{B}_1$ 中采样, 使用尽可能多的经验快速习得大致的策略参数. 经过一段时间的训练后, 策略将积累足够多的高价值的经验, 将仅使用高价值的经验进行更新网络, 实现显著优于基准策略高效学习. 此外, 本文采用目标网络技术以增加学习过程的稳定性. 特别的, 由于 critic 的同步更新, 其输出的评价结果极易发散, 使得策略难以收敛. 因此, 通过引入同结构目标网络 π' 、V'_1 和 V'_2, 通过软更新 $\alpha'_i = (1-\varepsilon)\alpha'_i + \varepsilon\alpha_i,\ \ i = 1,\ 2,\ 3$ 提高训练的稳定性, 其中 $\varepsilon \in \mathbb{R}$ 表示正参数.

与此同时, 由于采用了两个独立的 critic 网络, 进一步使用它们中的保守估计更新 critic 目标: $y_i = r_i + \gamma \min_{j=1,\ 2} V'_j\,(\boldsymbol{s}_{i+1},\ \pi'(\boldsymbol{s}_{i+1}) + \boldsymbol{\epsilon}_1)$, 其中 i 表示数据序列下标, 并引入随机噪声 $\boldsymbol{\epsilon}_1 \sim \mathcal{N}(0,\ \sigma_2)$ 来平滑 actor 策略输出结果, 提高 critic 评价精度. 此外, 文中采用延迟策略更新技术, 即 critic 更新 $d \in \mathbb{R}$ 次后, 更新一次目标网络和 actor 策略, 以减少时间差分误差. 在此基础上, 无人机 i 的策略参数更新如下: $\alpha_1(t+1) = \alpha_1(t) + \lambda_{\alpha_1}\nabla_{\alpha_1}J(\pi)$, $\alpha_2(t+1) = \alpha_2(t) + \lambda_{\alpha_2}\nabla_{\alpha_2}Loss_1$, $\alpha_3(t+1) = \alpha_3(t) + \lambda_{\alpha_3}\nabla_{\alpha_3}\times Loss_2$, 其中 $J(\pi) = \mathbb{E}[R_1 \mid \pi]$ 为目标函数, λ_{α_1}, λ_{α_2}, $\lambda_{\alpha_3} \in \mathbb{R}$ 表示正的学习率. $Loss_j = \frac{1}{N_m}\sum_{i=1}^{N_m}[y_i - V_j \times (\boldsymbol{s}_i, \boldsymbol{a}_i)]^2$, $j = 1,\ 2$ 为双 critic 的损失函数, $N_m \in \mathbb{R}$ 代表每次从缓冲区采样的数据量大小.

2.3 引导策略协同探索分析

本小节将从理论与实际结合的角度进行性能分析, 说明所提性能函数引导学习方法的有效性. 与纯模型驱动的控制方法不同, 深度强化学习通过试错交互来探索最优策略, 通常无法获得严格的稳定性. 尽管如此, 我们还是对提出的引导式学习框架进行了深入分析. 本文的主要贡献是设计了性能函数与深度强化学习算法协同的控制策略. 集群控制系统的协同探索与稳定性分析如下. 首先, 分析性能函数驱动的集群控制引导策略的有效性. 其次, 在引导策略的基础上进行引导式学习, 通过双 critic 框架筛选出高价值的探索经验, 从而实现高效的学习策略更新. 最后, 通过引导策略有效的协同与合作, 保证了整个系统的稳定性.

为此, 首先分析性能函数驱动的集群控制引导策略的稳定性, 提出如下定理.

定理 1. 针对无人机集群系统 (1), 当给定初始状态满足 $-\beta_{ij} < e_{pij}(0) < \beta_{ij}$, $i = 0, \cdots, n$, $j = x, y, z$, 在引导策略 (11)、(12) 的作用下, 集群系统状态收敛到平衡点 $\lim_{t\to\infty} \boldsymbol{e}_{pi} = 0$, 且始终满足飞行约束 (3).

证明 1. 定义 Lyapunov 候选函数如下:

$$V_i = V_{pi} + V_{oi} \tag{17}$$

其中, V_{pi} 与 V_{oi} 分别表示外环位置子系统和内环姿态子系统的 Lyapunov 候选函数:

$$V_{pi} = \frac{m_i}{2} \sum_{j=x,\, y,\, z} \frac{k_{\varphi ij} e_{pij}^2}{\beta_{ij}^2 - e_{pij}^2} + \frac{1}{2}\|\boldsymbol{\zeta}_i\|^2 \tag{18}$$

$$V_{oi} = \frac{1}{2}\boldsymbol{\eta}_i^\top \mathbb{I}_i \boldsymbol{\eta}_i + k_i \Psi_i \tag{19}$$

其中 $\Psi_i = \frac{1}{2}\mathrm{tr}\left[I - R_{di}^\top R_i\right]$ 表示定义在 SO(3) 流型上的误差函数, 其值始终满足 $\|\boldsymbol{e}_{Ri}\|^2/2 \le \Psi_i \le k_{\Psi i}\|\boldsymbol{e}_{Ri}\|^2$, 且 $k_{\Psi i} \in \mathbb{R}$ 为正的参数, 可求得其导数为 $\dot{\Psi}_i = \boldsymbol{e}_{Ri}\boldsymbol{e}_{\Omega i}$. 所提 Lyapunov 候选函数 (17) 为非负函数, 其值满足:

$$0 \le V_i \le \lambda_{\max}\left(\sum_{j=x,\, y,\, z} \frac{e_{pij}^2}{\beta_{ij}^2 - e_{pij}^2} + \|\boldsymbol{\xi}_i\|^2\right) \tag{20}$$

其中, $\lambda_{\max} = \max\{m_i k_{\varphi ij}/2,\ 1/2,\ \mathbb{I}_i/2,\ k_i k_{\Psi i}\}$ 表示其中系数的最大值, $\boldsymbol{\xi}_i = [\|\boldsymbol{\zeta}_i\| \quad \|\boldsymbol{\eta}_i\| \quad \|\boldsymbol{e}_{Ri}\|]^\top \in \mathbb{R}^3$ 为辅助变量. 由于全驱内环子系统指数收敛, 可将耦合变量 $\boldsymbol{d}_i \approx 0$, 进而对 Lyapunov 候选函数求导, 可得:

$$\begin{aligned}\dot{V}_i = &-\sum_{j=x,\, y,\, z} \frac{k_{\varphi ij} K_{pij} \beta_{ij}^2 e_{pij}^2}{(\beta_{ij}^2 - e_{pij}^2)^2} - K_{\zeta i}\boldsymbol{\zeta}_i^\top \boldsymbol{\zeta}_i \\ &- k_{\eta i}\boldsymbol{\eta}_i^\top \boldsymbol{\eta}_i - k_i k_{Ri} \boldsymbol{e}_{Ri}^\top \boldsymbol{e}_{Ri}\end{aligned} \tag{21}$$

考虑到在飞行误差约束内, 满足 $\beta_{ij}^2/(\beta_{ij}^2 - e_{pij}^2)^2 \ge 1$, 其值可进一步缩放为:

$$\dot{V}_i \le -\lambda_{\min}\left(\sum_{j=x,\, y,\, z} \frac{e_{pij}^2}{\beta_{ij}^2 - e_{pij}^2} + \|\boldsymbol{\xi}_i\|^2\right) \tag{22}$$

其中, $\lambda_{\min} = \min\{k_{\varphi ij}K_{pij},\ K_{\zeta i},\ k_{\eta i},\ k_i k_{Ri}\}$ 表示其中系数的最小值. 进而, 将 (20) 代入 (22) 可得 $\dot{V}_i \le -KV_i$, 其中 $K = \lambda_{\min}/\lambda_{\max} \in \mathbb{R}$ 为正的参数. 至此, 系统指数收敛得证. 进而, 可推导得出集群系统中无人机 i 的飞行误差始终满足:

$$-\mu\beta_{ij} \le e_{pij} \le \mu\beta_{ij}, \quad j = x,\ y,\ z \tag{23}$$

其中 $\mu = \sqrt{\frac{2V_i(0)}{2V_i(0) + m_i k_{\varphi ij}}} < 1$ 为正的参数.

在引导策略收敛的基础上, 通过设计双 critic 架构, 实现了对探索动作优劣的准确判断. 并且使用优于引导策略的探索经验更新网络, 解决了由随机探索策略和复杂动力学模型等导致的探索失败问题. 通过引导学习, 性能函数辅助深度强化学习探索更好的策略, 从而保证了高效的学习和可靠的泛化能力.

3 实验与结果分析

本节将设计集群飞行实验, 验证所提方法的有效性. 首先, 搭建仿真训练环境, 部署训练算法 1, 在经过训练后, 将所得到的深度强化学习策略应用于无人机集群控制. 具体而言, 本节进行了 3 组实验测试. (1) 在实验 1 中设计直线路径与领航-跟随编队队形, 采用所提策略控制无人机集群完成编队飞行任务, 验证所提策略的有效性. (2) 在实验 2 中设计更复杂的圆形轨迹, 并同时改变集群内的领航-跟随编队拓扑结构, 测试所提策略的泛化能力. (3) 实验 3 引入 x, y, z 三轴速度分量均为 0.5m/s 的风干扰, 同时保持目标轨迹、编队拓扑等条件与实验 (2) 一致, 验证所提策略的鲁棒性.

3.1 策略训练

本文基于 Ubuntu20.04 系统开发了无人机集群训练环境, 如图 4 所示. 具体而言, 基于 ROS 设计了无人机集群通讯框架, 其中的无人机模型使用 RotorS[32] 无人机仿真库中的 Firefly 六旋翼无人机和 Pelican 四旋翼无人机. 使用 PyTorch 搭建深度强化学习网络, 并在 Gazebo 软件中进行动力学交互训练. 完成训练后, 将得到的控制策略直接应用于无人机集群控制测试. 本文的训练参数如表 1 所示.

3.2 实验 1

为了充分测试所提方法的有效性, 采用 Firefly 与 Pelican 两种类型的旋翼无人机组成异构无人机集群. 在实验 1 中, 采用 1 架 Firefly 六旋翼无人机 U_1 和 4 架 Pelican 四旋翼无人机 $U_{2,\,3,\,4,\,5}$ 组成领航-跟随编队, 其拓扑结构如图 5(a) 所示. 设计虚拟领航者的飞行路径为 $x_0 = y_0 = 1.273\,2\ s$ m, $z_0 = 0.636\,6\ s + 2$ m, 其中 $s \in (0,\ 2\pi)$ 为路径坐标, 在此基础上, 采用时间最优轨迹规划算法计算实时轨迹[33], 将期望的实时轨迹输入所提控制算法, 驱动无人机集群完成飞行任务. 实验结果如图所示, 从图 4 中可以看出, 所提学习策略可以高效地控制集群无人机系统完成编队飞行任务. 具体而言, 从图 5(b) 中

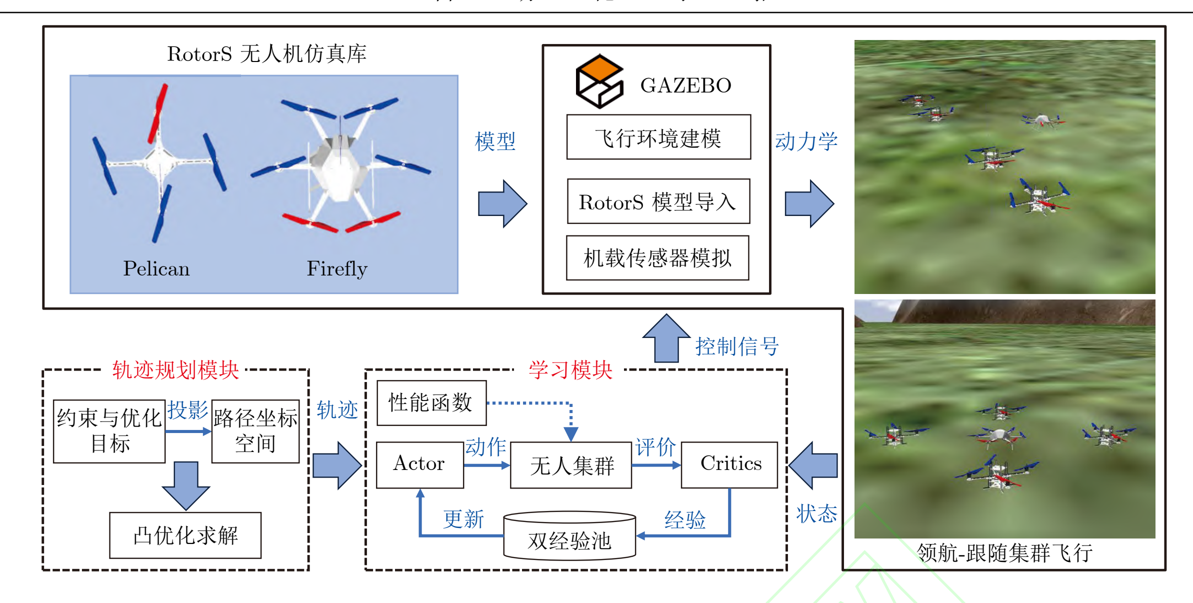


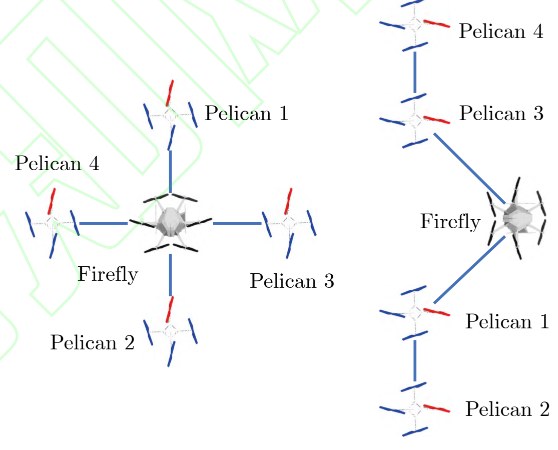
图 4 无人机集群控制策略训练与测试框架

Fig.4 Uav cluster control strategy training and testing framework

表 1

Table 1

参数	值
无人机质量 m_i	$m_1 = 1.6\text{kg},\ m_{2-5} = 1.0\text{kg}$
无人机转动惯量 $\mathbb{I}_i$	$\text{diag}[0.01\ \ 0.01\ \ 0.01]\text{kg}\cdot\text{m}^2$
重力加速度	9.8m/s^2
学习率 $\lambda_{\alpha_{1,\ 2,\ 3}}$	$1,\ 2,\ 2\times10^{-4}$
训练回合数 $M_{\max}$	100
训练步数 $N_{\max}$	500
经验池大小 $\mathcal{B}_{1,\ 2}$	10000
采样数据量 N_m	128
训练折扣因子 γ	0.95
探索与平滑系数 $\sigma_{1,\ 2}$	0.1, 0.05
控制策略交互频率	100 Hz
引导策略参数 $k_{\varphi ij}$, β_{ij}	0.2, 0.3
辅助增益矩阵 K_{pi}	diag[4 4 4]
辅助增益矩阵 K_{Ri}	diag[1.5 1.5 1.5]
外环控制参数 $K_{\zeta i}$	diag[2 2 2]
内环控制参数 $k_{\eta i}$	diag[1.5 1.5 1.5]
内环控制参数 k_i	diag[2 2 2]



(a) 实验 1 中的集群拓扑结构 (b) 实验 2 中的集群拓扑结构

图 5 无人机集群拓扑结构

Fig.5 Topology of UAV cluster

可以发现, 无人机集群中的 5 架异构无人机的飞行误差都被约束在公式 (3) 中, 集群飞行误差远远小于预设的 0.3 m 约束. 此外, 从图 5(a) 中可以看出, 无人机集群中每个无人机都能准确地跟踪期望速度, 完成编队飞行任务. 值得一提的是, 编队飞行的速度存在部分抖震现象, 这是由于无人机中未建模动动力学因素的影响, 特别是集群飞行速度较大时, 跟随子节点的无人机难以复刻所有的领航节点动作, 导致速度抖震. 为此, 提出了策略引导的深度强化学习无人机集群控制方法, 有效抑制无人机集群的飞行误差, 实现精准集群控制.

3.3 实验 2

为了充分验证所提方法, 在实验 1 的基础上, 进一步采用直径为 10 m 的圆形期望路径: $x_0 = -5\sin(s)$ m, $y_0 = -5\cos(s) + 5$ m, $z_0 = -\cos(s) + 3$m, 其中 $s \in (0,\ 2\pi)$ 为路径坐标, 其编队拓扑连接图如图 3(b) 所示. 使用所提方法控制无人机集群完成上述圆形轨迹飞行任务, 并同时保持编队队形. 实验结果如图 6 ~ 9 所示. 从图 6 和 7 可见, 所提方法能够有效地控制无人机集群沿给定轨迹飞行, 在

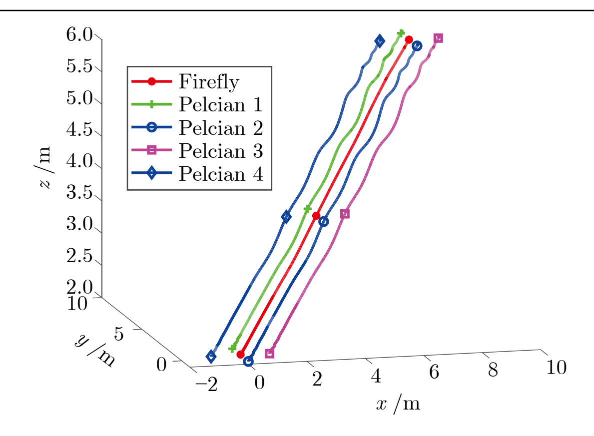


图 6　无人机集群的编队飞行轨迹

Fig. 6　Drone cluster formation flight trajectory

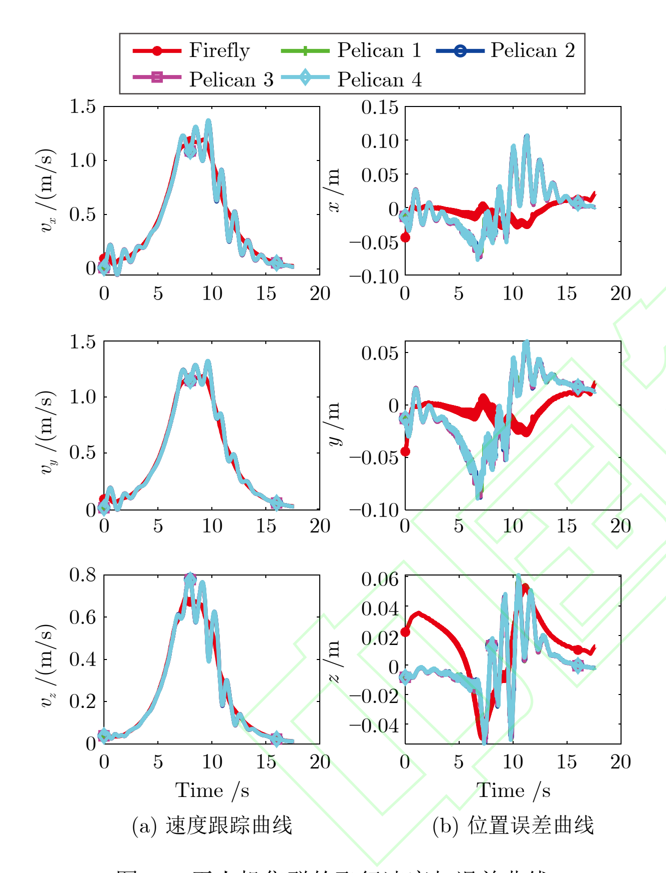


图 7　无人机集群的飞行速度与误差曲线

Fig. 7　Flight speed and error curve of UAV cluster

x, y, z 三个方向都能以期望队形完成复杂曲线飞行. 此外, 通过分析飞行过程中的集群误差图 8, 可以看出所有无人机在 x, y, z 三个方向的飞行误差都得到有效抑制, 其值远小于预设误差约束.

为了充分说明所提引导策略的有效性, 本文记录了实验 1 和实验 2 在集群飞行过程中每个无人机控制策略中两个 critic 的比较输出, 即Q_1, Q_2 , 分别如图 11(a) 和图 11(b) 所示. 可见, 在集群飞行任务中, 始终满足 $Q = \min\{Q_1, Q_2\} > 0$, 表明所提的深度强化学习策略输出动作评分优于引导策略, 能够有效地控制无人机集群完成飞行任务.

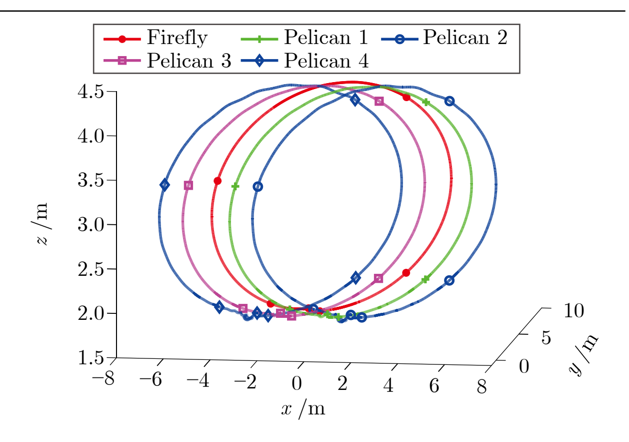


图 8　无人机集群的飞行轨迹曲线

Fig. 8　Uav cluster flight trajectory curve

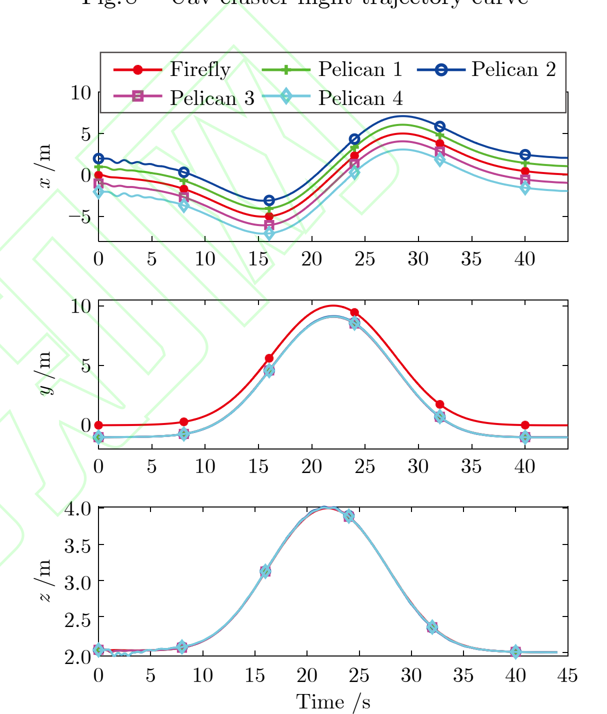


图 9　无人机集群在 x, y, z 三个方向的飞行曲线

Fig. 9　The flight curve of the drone cluster in x , y and z directions

3.4　实验 3

为了进一步验证所提方法的鲁棒性, 分别在 x, y, z 方向上加入风力大小为0.5 m/s 的恒风干扰, 其他实验条件均与实验 2 相同. 实验结果如图 12 所示, 所提方法虽然在有风干扰下轨迹会产生细微波动, 但仍能够有效地控制无人机集群沿给定轨迹飞行, 并在 x, y, z 三个方向都能以期望队形完成复杂曲线飞行. 此外通过分析集群误差图 13(a)

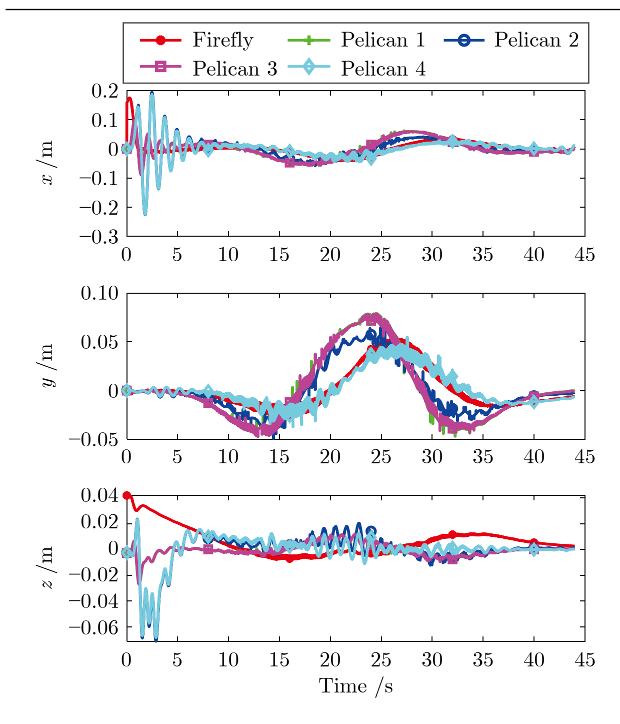


图 10 无人机集群在 x, y, z 三个方向的飞行误差

Fig. 10 Flight error of UAV cluster in x, y and z directions

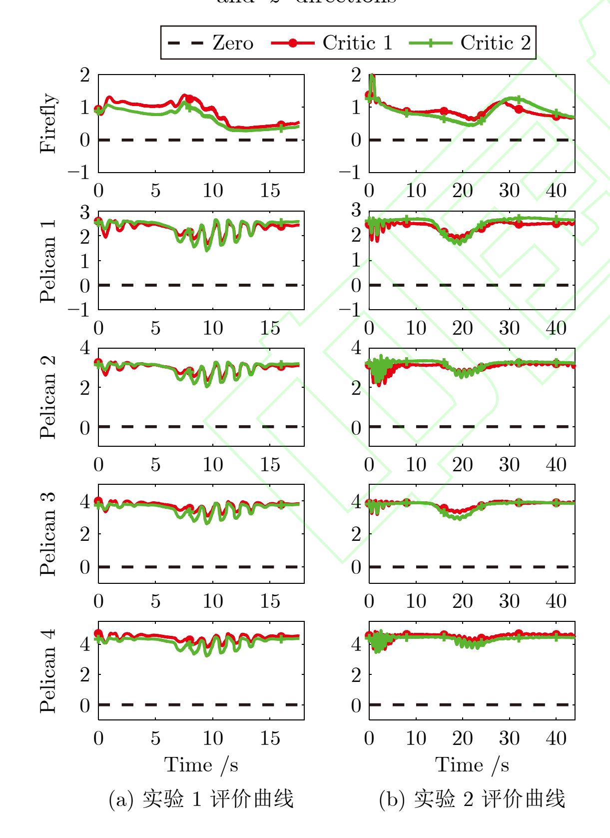


图 11 深度强化学习策略与引导策略在双 critic 框架中的评价曲线

Fig. 11 Evaluation curves of deep reinforcement learning strategies and guidance strategies in the dual critic framework

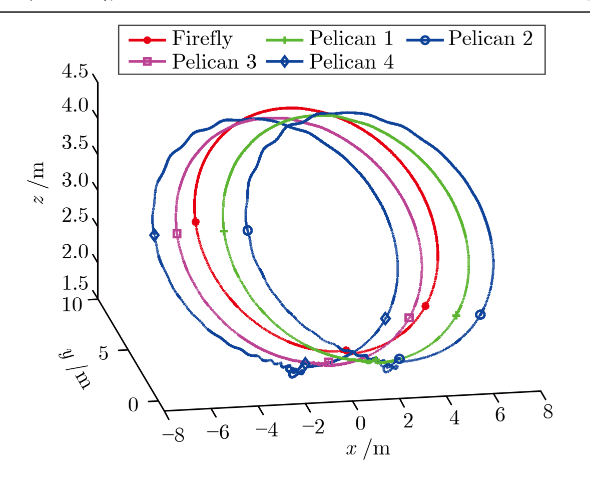


图 12 无人机集群的飞行轨迹曲线

Fig. 12 The flight trajectory curves of drone swarms

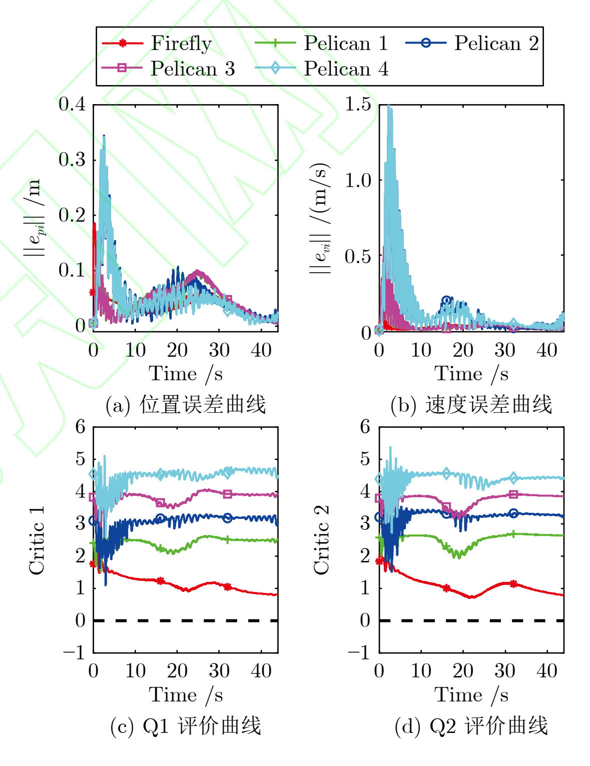


图 13 无人机集群的误差与评价曲线

Fig. 13 Error and Evaluation Curves of Drone Swarms

和 13(b), 以及评价曲线图 13(c) 和 13(d) 可得, 在飞行过程中无人机集群的位置误差与速度误差均得到了有效抑制, 且从评价曲线可以看出 $Q = \min\{Q_1, Q_2\} > 0$, 表明本文所提深度强化学习算法有很好的鲁棒性, 能够有效的完成无人机集群的飞行控制任务.

综上所述, 本文将深度强化学习方法和性能函数结合, 提出了性能函数引导的深度强化学习控制方法, 保证学习策略优于引导控制策略, 有效提高了无人机编队控制精度.

4 结论

针对无人机集群控制问题, 本文提出了一种性能函数引导的深度强化学习控制方法, 形成的模型-数据混合驱动的控制策略, 保证了高性能集群控制. 首先, 在领航-跟随集群框架基础上, 设计性能函数驱动的集群控制方法, 将集群误差约束在预期范围内. 在此基础上, 设计了性能函数引导的深度强化学习控制策略, 利用性能函数的示范经验, 辅助训练学习网络, 解决了性能函数在复杂工况下极易失效的安全隐患, 同时极大提升了学习策略的训练效率与可靠性. 将无人机系统的先验模型与深度强化学习策略结合, 设计的性能函数驱动的学习控制策略, 充分利用了非线性控制与深度强化学习的优势, 增强了无人机集群的鲁棒性, 提高了飞行精度. 最后, 将所提出的算法部署在 ROS 无人机集群平台上, 实验结果验证了性能函数驱动的深度强化学习算法的有效性.

References

1 Chen M, Ma H X, Yong K N, Wu Y. Safety Flight Control of UAV:A Survey. *Robotics*, 2023, **45**: 345−366
(陈谋, 马浩翔, 雍可南, 吴颖. 无人机安全飞行控制综述. 机器人, 2023, **45**: 345−366)

2 Erskine J, Briot S, Fantoni I, Chriette A. Singularity Analysis of Rigid Directed Bearing Graphs for Quadrotor Formations. *IEEE Transactions on Robotics*, 2024, **40**: 139−157

3 Dai B, He Y Q, Gu F, Wang Q H, Xu W L. Acceleration Feedback Enhanced Controller for Wind Disturbance Rejection of Rotor Unmanned Aerial Vehicle. *Robotics*, 2020, **42**(1): 79−88
(代波, 何玉庆, 谷丰, 王骞翰, 徐卫良. 基于加速度反馈增强的旋翼无人机抗风扰控制. 机器人, 2020, **42**(1): 79−88)

4 Cai Y S, Xu J, Niu Y G. Attitude Consensus Control of UAV Swarm Based on Adaptive Multi-scale Super-twisting Algorithm. *Acta Automatica Sinica*, 2023, **49**: 1656−1666
(蔡运颂, 许璟, 牛玉刚. 基于自适应多尺度超螺旋算法的无人机集群姿态同步控制. 自动化学报, 2023, **49**: 1656−1666)

5 Ille M, Namerikawa T. Collision Avoidance between Multi-UAV-Systems Considering Formation Control using MPC. 2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM), 2017: 651−656

6 Ru C J, Wei R X, Dai J, Shen D, Zhang L P. Autonomous Reconfiguration Control Method for UAV s Formation Based on Nash Bargain. *Acta Automatica Sinica*, 2013, **39**: 1349−1359
(茹常剑, 魏瑞轩, 戴静, 沈东, 张立鹏. 基于纳什议价的无人机编队自主重构控制方法. 自动化学报, 2013, **39**: 1349−1359)

7 Qi J T, Guo J J, Wang M M, Wu C, Ma Z W. Formation Tracking and Obstacle Avoidance for Multiple Quadrotors With Static and Dynamic Obstacles. *IEEE Robotics and Automation Letters*, 2022, **7**(2): 1713−1720

8 Shi Y, Hua Y Z, Yu J L, Dong X W, Lü J H, Ren Z. Cooperative Fault-Tolerant Formation Tracking Control for Heterogeneous Air-Ground Systems Using a Learning-Based Method. *IEEE Transactions on Aerospace and Electronic Systems*, 2024, **60**(2): 1505−1518

9 Zhang Y, Ma L, Yang C Y, Zhou L N, Wang G Q, Dai W. Formation Control for Multiple Quadrotors Under DoS Attacks via Singular Perturbation. *IEEE Transactions on Aerospace and Electronic Systems*, 2023, **59**(4): 4753−4762

10 Park B S, Yoo S J. Time-Varying Formation Control With Moving Obstacle Avoidance for Input-Saturated Quadrotors With External Disturbances. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2024, **54**(5): 3270−3282

11 Du H B, Zhu W W, Wen G H, Duan Z S, Lü J H. Distributed Formation Control of Multiple Quadrotor Aircraft Based on Nonsmooth Consensus Algorithms. *IEEE Transactions on Cybernetics*, 2019, **49**(1): 342−353

12 Dong X W, Yu B C, Shi Z Y, Zhong Y S. Time-Varying Formation Control for Unmanned Aerial Vehicles: Theories and Applications. *IEEE Transactions on Control Systems Technology*, 2015, **23**(1): 340−348

13 Hu Z J, Jin X. Formation Control for an UAV Team With Environment-Aware Dynamic Constraints. *IEEE Transactions on Intelligent Vehicles*, 2024, **9**(1): 1465−1480

14 Wang Z X, Zou Y, Liu Y Z, Meng Z Y. Distributed Control Algorithm for Leader-Follower Formation Tracking of Multiple Quadrotors: Theory and Experiment. *IEEE/ASME Transactions on Mechatronics*, 2021, **26**(2): 1095−1105

15 Liu H, Ma T, Lewis F L, Wan Y. Robust Formation Trajectory Tracking Control for Multiple Quadrotors With Communication Delays. *IEEE Transactions on Control Systems Technology*, 2020, **28**(6): 2633−2640

16 Wu J, Luo C, Min G, McClean S. Formation Control Algorithms for Multi-UAV Systems with Unstable Topologies and Hybrid Delays. *IEEE Transactions on Vehicular Technology*, 20241−12

17 Dai S L, He S D, Chen X, Jin X. Adaptive Leader-Follower Formation Control of Nonholonomic Mobile Robots With Prescribed Transient and Steady-State Performance. *IEEE Transactions on Industrial Informatics*, 2020, **16**(6): 3662−3671

18 Shen Y Y, Zhou J, Xu Z D, Zhao F G, Xu J M, Chen J M, Li S. Aggressive Trajectory Generation for a Swarm of Autonomous Racing Drones. *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 20237436−7441

19 Song F L, Li Z, Yu X H. A Feedforward Quadrotor Disturbance Rejection Method for Visually Identified Gust Sources Based on Transfer Reinforcement Learning. *IEEE Transactions on Aerospace and Electronic Systems*, 2023, **59**(5): 6612−6623

20 Xiao C X, Lu P, He Q X. Flying Through a Narrow Gap Using End-to-End Deep Reinforcement Learning Augmented With Curriculum Learning and Sim2Real. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, **34**(5): 2701−2708

21 Han H R, Cheng J, Xi Z L, Yao Bingcai. Cascade Flight Control of Quadrotors Based on Deep Reinforcement Learning. *IEEE Robotics and Automation Letters*, 2022, **7**(4): 11134−11141

22 Hua H A, Fang Y C. A Novel Reinforcement Learning-Based Robust Control Strategy for a Quadrotor. *IEEE Transactions on Industrial Electronics*, 2023, **70**(3): 2812−2821

23 Zhao W B, Liu H, Lewis F L. Robust Formation Control for Cooperative Underactuated Quadrotors via Reinforcement Learning. *IEEE Trans. Neural Netw. Learn. Syst.*, 2021, **32**(10): 4577−4587

24 Hua H A, Fang Y C. A Novel Learning-Based Trajectory Generation Strategy for a Quadrotor. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, **35**(7): 9068−9079

25 Hwangbo J, Sa I, Siegwart R, Hutter M. Control of a Quadrotor With Reinforcement Learning. *IEEE Rob. Autom. Lett.*, 2017, **2**(4): 2096−2103

26 Pu Z Q, Wang H M, Liu Z, Yi J Q, Wu S G. Attention Enhanced Reinforcement Learning for Multi-agent Cooperation. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, **34**(11): 8235−8249

27 Sun Q Y, Fang J B, Zheng W X, Tang Y. Aggressive Quadrotor Flight Using Curiosity-Driven Reinforcement Learning. *IEEE Transactions on Industrial Electronics*, 2022, **69**(12):

13838–13848

28 Wang Y D, Sun J, He H B, Sun C Y. Deterministic Policy Gradient With Integral Compensator for Robust Quadrotor Control. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, **50**(10): 3713–3725

29 Raja G, Essaky S, Ganapathisubramaniyan A, Baskar Y. Nexus of Deep Reinforcement Learning and Leader-Follower Approach for AIoT Enabled Aerial Networks. *IEEE Transactions on Industrial Informatics*, 2023, **19**(8): 9165–9172

30 Yoo J, Jang D, Kim H J, Johansson K H. Hybrid Reinforcement Learning Control for a Micro Quadrotor Flight. *IEEE Control Systems Letters*, 2021, **5**(2): 505–510

31 Koryakovskiy I, Kudruss M, Vallery H, Babuska R, Caarls W. Model-Plant Mismatch Compensation Using Reinforcement Learning. *IEEE Robotics and Automation Letters*, 2018, **3**(3): 2471–2477

32 Furrer F, Burri M, Achtelik M, and Siegwart R, RotorS-A Modular Gazebo MAV Simulator Framework. Springer International Publishing, 2016, pp. 595-625.

33 Hua H, Fang Y, Zhang X, and Qian C. A time-optimal trajectory planning strategy for an aircraft with a suspended payload via optimization and learning approaches. *IEEE Transactions on Control Systems Technology*, 2022, **30**(6): 2333–2343



王耀南 中国工程院院士, 湖南大学电气与信息工程学院教授. 主要研究方向为机器人学, 智能控制和图像处理. E-mail: yaonan@hnu.edu.cn
(**WANG Yao-Nan** Academician at Chinese Academy of Engineering, professor at the College of Electrical and Information Engineering, Hunan University. His research interest covers robotics, intelligent control, and image processing.)



华和安 湖南大学电气与信息工程学院副研究员. 主要研究方向为空中机器人的智能规划、控制与集群.
E-mail: huahean@hnu.edu.cn
(**HUA He-An** Associate researcher atthe College of Electrical and Information Engineering, Hunan University. His main research interest is intelligent planning, control, and swarming of aerial robots.)



张 辉 湖南大学机器人学院教授. 主要研究方向为机器视觉, 图像处理和机器人控制. 本文通信作者.
E-mail: zhanghuihby@126.com
(**ZHANG Hui** Professor at the School of Robotics, Hunan University. His research interest covers machine vision, image processing, and robot control. Corresponding author of this paper.)



钟 杭 湖南大学机器人学院副教授. 主要研究方向为机器人控制, 视觉伺服和路径规划.
E-mail: zhonghang@hnu.edu.cn
(**ZHONG Hang** Associate Professor at the School of Robotics, Hunan University. His research interest covers robot control, visual servoing, and path planning.)



樊叶心 湖南大学机器人学院博士后. 主要研究方向为 机器人感知与控制、深度强化学习及运动规划.
E-mail: yexinfan@hnu.edu.cn
(**FAN Ye-Xin** Postdoctoral fellow at the School of Robotics, Hunan University. Her research interests include robot perception and control, deep reinforcement learning and motion planning.)



梁鸿涛 湖南大学电气与信息工程学院博士研究生. 主要研究方向为空中机器人集群运动控制与路径规划.
E-mail: lianghongtao1@hnu.edu.cn
(**LIANG Hong-Tao** PhD candidate at the College of Electrical and Information Engineering, Hunan University. His research interests include swarm motion control and path planning for aerial robots.)



常 浩 湖南大学电气与信息工程学院博士研究生. 主要研究方向为空中机器人的视觉感知与路径规划.
E-mail: changhao@hnu.edu.cn
(**CHANG Hao** PhD candidate at the College of Electrical and Information Engineering, Hunan University. His research interests include visual perception and path planning for aerial robots.)



方勇纯 南开大学机器人与信息自动化研究所教授. 主要研究方向为非线性控制、视觉伺服、欠驱动系统控制和基于原子力显微镜的纳米系统.
E-mail: fangyc@nankai.edu.cn
(**FANG Yong-Chun** Professor at the Institute of Robotics and Automatic Information Systems, Nankai University. His research interest covers nonlinear control, robot visual servoing control, control of underactuated systems and AFM-based nanosystems.)